

The Complete Amino Acid Sequence of Three Alcohol Dehydrogenase Alleloenzymes (Adh^{N-11} , Adh^S and Adh^{UF}) from the Fruitfly *Drosophila melanogaster*

David R. THATCHER

Department of Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3HR, Scotland, U.K.

(Received 27 November 1979)

The sequence of three alcohol dehydrogenase alleloenzymes from the fruitfly *Drosophila melanogaster* has been determined by the sequencing of peptides produced by trypsin, chymotrypsin, thermolysin, pepsin and *Staphylococcus aureus*-V8-proteinase digestion. The amino acid sequence shows no obvious homology with the published sequences of the horse liver and yeast enzymes, and secondary structure prediction suggests that the nucleotide-binding domain is located in the *N*-terminal half of the molecule. The amino acid substitutions between Adh^{N-11} (a point mutation of Adh^F), Adh^S and Adh^{UF} alleloenzymes were identified. Adh^{N-11} alcohol dehydrogenase differed from the other two by a glycine-14- (Adh^S and Adh^{UF})-to-aspartic acid substitution, the Adh^S enzyme from Adh^{N-11} and Adh^{UF} enzymes by a threonine-192- (Adh^{N-11} and Adh^{UF})-to-lysine (Adh^S) substitution and the Adh^{UF} enzyme was found to differ by an alanine-45- (Adh^S and Adh^{N-11})-to-aspartic acid (Adh^{UF}) charge substitution and a 'silent' asparagine-8- (Adh^S and Adh^{N-11})-to-alanine (Adh^{UF}) substitution. Detailed sequence evidence has been deposited as Supplementary Publication SUP 50107 (36 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1978) 169, 5.

The fruitfly *Drosophila melanogaster* produces large amounts of a low-molecular-weight dimeric alcohol dehydrogenase (EC 1.1.1.1) that apparently lacks a requirement for a metal cofactor. This enzyme continues to be the object of much study both in population and molecular biology. A large amount of information has now accumulated on the amount and distribution of allelic variation at the alcohol dehydrogenase (*Adh*) locus (Vigue & Johnson, 1973; David, 1977) and mechanistic theories accounting for this variation have received some experimental support (Clarke, 1975). The *Adh* locus has been mapped cytogenetically (O'Donnell *et al.*, 1977; Woodruff & Ashburner, 1979*a,b*), and its genetic fine structure has been investigated by the analysis of allelic recombination and by the use of DNA-cloning methods.

The present paper describes the primary protein sequence of three allelic variant enzymes in *D. melanogaster*. Partial sequences of this enzyme have been published by Schwartz & Jornvall (1976), Thatcher & Camfield (1977), Fletcher *et al.* (1978),

Abbreviation used: SDS, sodium dodecyl sulphate.

Auffret *et al.* (1978), and Retzios & Thatcher (1979).

Experimental

Organisms

Isogenic strains of Adh^{N-11} and Adh^S were obtained from Professor B. Clarke (Genetics Research Unit, University of Nottingham, Nottingham, U.K.). Strain Adh^{UF} was obtained by Dr. D. A. Briscoe (Department of Genetics, University of Edinburgh, Edinburgh, Scotland, U.K.). Flies were maintained and grown in bulk by using large-population cages as described by Thatcher (1977). Cages were used for egg laying only and larvae were cultured in plastic boxes 20 cm × 32 cm × 7 cm on a cornmeal/molasses medium. Larvae were collected at the third-instar stage of development by floating in 14% (w/v) sucrose solutions. Only one strain was grown in bulk at any one time and stocks were maintained in a separate room to avoid contamination.

Chemicals

Pepsin A (EC 3.4.23.1; twice-crystallized), trypsin (EC 3.4.21.4; TRSF-6188) and chymotrypsin (EC 3.4.21.1; thrice-crystallized; CD 160-9) were obtained from the Worthington Biochemical Corporation, Freehold, NJ, U.S.A. Soya-bean trypsin inhibitor was purchased from Seravac Laboratories, Colnbrook, Bucks., U.K. Thermolysin (EC 3.4.24.4) was obtained from Chugai Boyaki Co., P.O. Box 106, Higashi, Osaka, Japan. Carboxypeptidase A (EC 3.4.17.1; di-isopropyl phosphorofluoridate-treated) was purchased from Sigma (London) Chemical Co., Kingston-upon-Thames, Surrey, U.K. Leucine aminopeptidase (EC 3.4.11.1) was bought from Boehringer, Mannheim, Germany. Staphylococcal proteinase V8 (EC 3.4.12.—) was purified by a modification of the methods of Drapeau *et al.* (1972) and Ryden *et al.* (1974). Toluene-*p*-sulphonic acid and phenyl isothiocyanate were obtained from the Pierce Chemical Co., P.O. Box 117, Rockford, IL, U.S.A. Sephadex grades, Sepharose 4B and Pharmalyte ampholytes were purchased from Pharmacia (G.B.) Ltd., Hounslow, Middx., U.K. Ion-exchange celluloses and chromatography paper were obtained from Whatman Biochemicals, Maidstone, Kent, U.K. Acrylamide monomer and *NN'*-methylenebisacrylamide were bought from the Eastman-Kodak Chemical Co., Rochester, NY, U.S.A. Thin-layer cellulose plates were obtained from E. Merck, Darmstadt, Germany, and fluorecamine was kindly given by Dr. J. Hermoso, Instituto de Biología, Madrid, Spain. All other laboratory reagents and biochemicals were obtained from BDH Ltd., Poole, Dorset, U.K.

Protein purification

The enzyme from strain *Adh*^{UF} was purified to homogeneity (on the basis of a single band on an SDS/polyacrylamide gradient gel) by the method described by Thatcher (1977). The enzyme was also prepared from strain-*Adh*^{N-11} flies by this method, the inactive enzyme being detected by SDS/polyacrylamide-gel electrophoresis. *Adh*^S enzyme was prepared by the following variation of the general method. The (NH₄)₂SO₄ fraction was prepared as described previously (Thatcher, 1977), then desalted into 0.01 M-Tris/HCl buffer, pH 7.4 by gel filtration on Sephadex G-25. The enzyme solution was then passed through a column [10 cm (diam.) × 50 cm] of DEAE-cellulose equilibrated with the same buffer. The crude eluate was then passed through a column [2.5 cm (diam.) × 10 cm] of CM-cellulose equilibrated with the Tris buffer. As the pI of the *Adh*^S enzyme is 7.45, the alcohol dehydrogenase passed through both ion-exchange columns without adsorption and emerged 90% pure as judged by SDS/polyacrylamide-gel electrophoresis of an oxidized sample. Final purification from a

contaminating proteolytic enzyme was achieved by repeated gel filtration with Sephadex G-100 as detailed by Thatcher (1977).

Isoelectric focusing

Isoelectric focusing was performed in thin-layer agarose gels (Agarose-EF; LKB, Bromma, Sweden) by using an LKB Multiphor apparatus as directed by the makers. The gel was prepared by boiling 0.264 g of sorbitol and 3.3 g of agarose in 30 ml of water for 10 min in a reflux apparatus. The gel solution was then cooled to 60°C and 2 ml of wide-range (pH 3–10) Pharmalytes added before pouring on to a clean glass plate. On setting, the gel was blotted with Whatman no. 3 chromatography paper for 10 min and the samples applied. Isoelectric focusing occurred after 3 h at 5 W constant power, when the coloured marker proteins, cytochrome *c*-551 and azurin, formed compact bands. The pH gradient was measured with a 3 mm²-area flat-ended pH-electrode (Russell pH Ltd., Auchtermuchty, Fife, Scotland, U.K.).

Sequencing methods

All the methods used were either described or cited in Thatcher (1975), Ambler & Wynn (1973) and Ambler (1973). Oxidized protein was digested with proteinases and the resulting peptides purified by Sephadex G-25 gel filtration, followed by high-voltage paper electrophoresis and chromatography. The *N*-terminal sequence of some of the large peptide fragments was investigated with an automatic sequencer (Beckman model 890A) by using a succinylated (3-carboxypropionylated) myoglobin carrier and the *NN*-dimethylbenzylamine double-cleavage program recommended by the makers. Amide residues were assigned from the electrophoretic mobilities of isolated peptides and by release of amino acids by leucine aminopeptidase. Peptides were also purified on t.l.c. plates by using the solvent systems described by Holder & Ingversen (1978). Many of the tryptic and staphylococcal-V8-proteinase-digested peptides were extremely hydrophobic and could not be separated by paper methods. These fragments were purified by chromatography on DEAE- and CM-cellulose columns in 6 M-urea buffers. Columns (1 cm diam. × 10 cm) of the ion-exchange celluloses were equilibrated with 0.02 M-ammonium acetate, pH 7.5 (DEAE-cellulose columns) or pH 6.0 (CM-cellulose columns) buffers. The peptide samples were diluted with the appropriate equilibration buffer and applied to the column. Peptides were eluted with a linear gradient of 0–0.15 M-NaCl and were located by their u.v. absorbance at 280 and 254 nm. After freeze-drying, the peptides were desalted into 5% (v/v) formic acid. This method produced peptide fractions of varying purity, and homogeneity was

established by amino-acid analysis and analytical peptide 'mapping' with chymotrypsin. Amino-acid analysis was performed with a Rank-Hilger J182 Chromaspek amino-acid analyser after hydrolysis in 6M-HCl or 3M-toluene-*p*-sulphonic acid.

A 1.5–2mmol sample of alcohol dehydrogenase was used in each proteinase digestion. Tryptic, thermolytic and chymotryptic digests were performed in 0.2M-ammonium acetate buffer, pH 8.5, at a proteinase/substrate ratio of 40:1 (w/w). Pepsin digestion was carried out in a 10% (v/v) aqueous solution of formic acid, at a protein/proteinase ratio of 80:1 (w/w). *Staphylococcus aureus* V8-proteinase digestions were performed in 0.2M-ammonium acetate buffer, pH 4.5, at an enzyme/substrate ratio of 1:20 (w/w). Digestions were at 37°C for 4h and were terminated by freeze-drying.

Results

Two major electrophoretic variants of alcohol dehydrogenase occur in natural populations of *Drosophila melanogaster*, namely the *Adh^F* or 'fast' enzyme, and *Adh^S* or 'slow' enzyme (Johnson & Dennison, 1964). Certain populations of *D. melanogaster* also contain low frequencies of three other electrophoretic types called 'ultra-fast' *Adh^{UF}*, 'ultra-slow' *Adh^{US}* and a slightly-faster-than-'fast' enzyme called *Adh^{F'}*. These differences in electrophoretic mobility are caused by marked differences in isoelectric point between the enzymes (Fig. 1). Each alleloenzyme is also associated with small amounts of two electronegative multiple forms that have been postulated as products of an interaction between the enzyme and NAD⁺ adducts (Schwartz & Sofer, 1976a) but could also be explained by deamidation of the major 'isoenzyme' (see below).

The complete amino-acid sequences of the enzyme produced by *Adh^{UF}*, *Adh^S*, and *Adh^{N-11}* strains has been determined. *Adh^{N-11}* is an ethylmethane-sulphonate-induced mutation of the *Adh^F* allele (O'Donnell *et al.*, 1977). Apart from the *Adh^{N-11}* mutant substitution, it is assumed that the primary sequence of the *Adh^{N-11}* enzyme is equivalent to that of *Adh^F*, the allele from which it was derived. A complete description of the quantitative results obtained in this work has been deposited as Supplementary Publication SUP 50107. The sequence of the *Adh^F* enzyme is presented in the one-letter code in Fig. 2 and a summary of the data used to deduce this sequence is shown in Fig. 3.

Reliability of the sequence

No one proteolytic digest covered the whole sequence in peptides. Tryptic peptides were isolated for the complete sequence except the region between arginine-103 and lysine-125, which was covered by staphylococcal-V8-proteinase-digested and chymo-

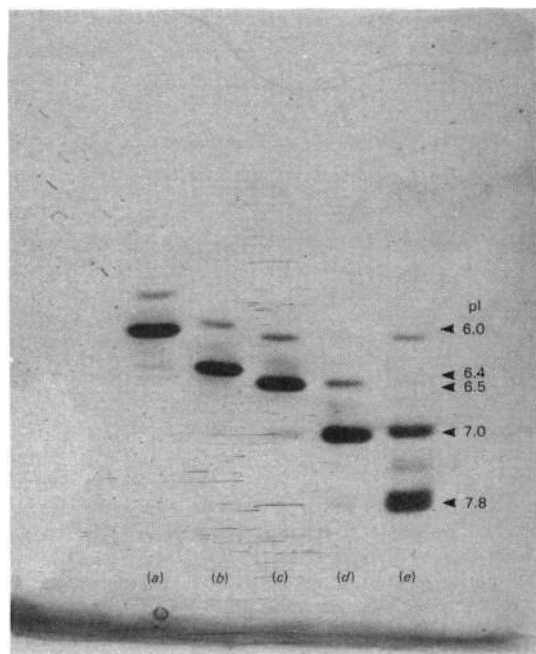


Fig. 1. Isoelectrofocusing of the *Adh* alleloenzymes. Isoelectric focusing was performed in an agarose-gel matrix as described in the Experimental section. Samples were prepared from flies isogenic for a particular allele and the focused gel was stained for alcohol dehydrogenase activity by the method of Grell *et al.* (1968). Track A was derived from strain *Adh^{UF}*, track B from strain *Adh^F*, track C from strain *Adh^{F'}*, track D from strain *Adh^S* and track E from strain *Adh^{US}*. pI values were deduced by the measurement of the surface pH of the gel before staining for enzyme activity.

tryptic peptides. Satisfactory overlaps were obtained for all residues in the sequence except phenylalanine-122. The main evidence that this residue is in fact overlapped comes largely from the close agreement between the composition of the proposed sequence and the analytical composition of the *Adh^{UF}* enzyme (Table 1) and from the composition of peptide T5a. The average percentage deviation of the sequence composition from the analytical composition is 3.7% for the residues stable to acid hydrolysis, rising to 6.1% when all residues are included in the comparison. Tryptophan-235 and -247 were identified unequivocally by isolation and analysis of thermolysin peptides H47 and H50a and by the analysis of fragments derived by BNPS-skatole [2-(2-nitrophenylsulphenyl)-3-methyl-3'-bromoindolenine] cleavage after tryptophan (D. R. Thatcher, unpublished work). Tryptophan-123 and

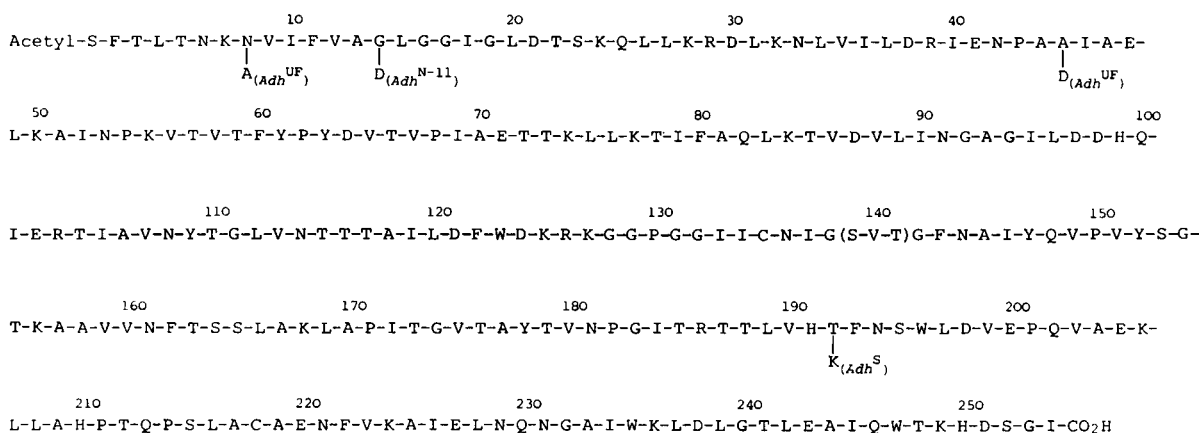


Fig. 2. Amino-acid sequence of the *Adh^F* alleloenzyme

The sequence is written in the one-letter code (see below) and the positions of the mutant substitutions are identified. The total sequence for enzymes *Adh^{N-11}*, *Adh^{UF}* and *Adh^S* was determined. One-letter code for amino acid sequences [for further details, see *Biochem. J.* (1969) 113, 1-4]: A, alanine; B, aspartic acid or asparagine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; X, unknown or 'other'; Y, tyrosine; Z, glutamic acid or glutamine.

Table 1. Amino-acid composition of the *Adh^{UF}* alcohol dehydrogenase alleloenzyme

Analytical data (column 1) was taken from Table 2 of Thatcher (1977). The number of residues/molecule of enzyme subunit was calculated from these data by using a divisor of 5.91, which gave the best fit for an enzyme of mol.wt. 27000. Column 4 shows the theoretical composition of the *Adh^F* alleloenzyme determined from Fig. 3.

Amino acid	Amount of amino acid recovered in analysis (μ mol)	Residues/molecule	Sequence composition
Asp	165.9	28.05	29
Thr	146.0	24.69	28
Ser	63.5	10.73	9
Glx	105.0	17.75	17
Pro	63.5	10.74	11
Gly	108.3	18.3	20
Ala	130.5	22.06	21
Val	125.5	21.2	22
Met	—	—	—
Ile	122.0	20.63	23
Leu	158.3	26.78	27
Tyr	36.1	6.12	6
Phe	52.9	8.94	9
His	23.5	4.02	4
Lys	105.8	17.89	17
Arg	32.2	5.4	5
Trp	27.1	4.85	4
Cys	18.6	3.14	2

-196 were assigned largely on the detection of intense u.v. fluorescence in the oxidized peptides covering this part of the sequence and by obvious gaps during dansyl-Edman degradation of peptides

C23 and H29. The termini of the enzyme molecule were identified unambiguously. The *N*-terminal blocking group of peptide T1 was shown to be acetyl by mass spectrometry (Auffret *et al.*, 1978), and

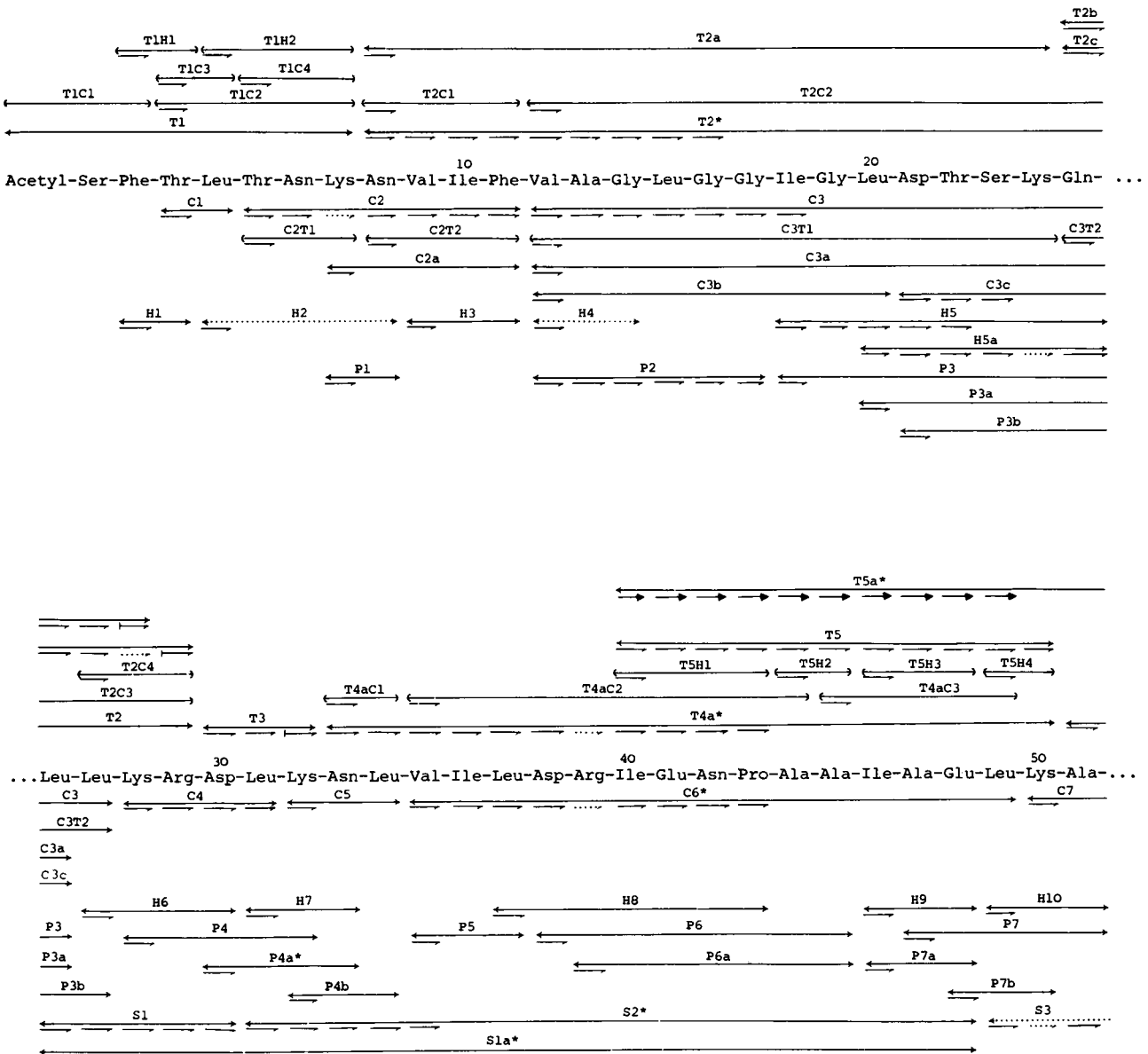


Fig. 3. Summary of the data used to deduce the sequence of the alcohol dehydrogenases

Peptides produced by digestion with trypsin (T), chymotrypsin (C), thermolysin (H), pepsin (P) or staphylococcal proteinase V8 (S) are denoted by \leftrightarrow . Peptides are numbered from the *N*-terminus and those from the same digest covering the same section of sequence are labelled by the major peptide number followed by a, b, etc. Subdigest peptide ($\leftarrow\rightarrow$) are named first by the number of the parent peptide and secondly by a letter denoting the subdigest proteinase e.g. T11bC2 is the second peptide obtained from tryptic peptide 11b by subdigestion with chymotrypsin. All peptides were analysed quantitatively for amino acid composition but those marked * were substandard and those marked $\leftrightarrow\leftrightarrow$ particularly bad. The dansyl-Edman method of manual peptide sequencing was represented by $\leftarrow\leftarrow$ if substandard). C-terminal amino acids released by carboxypeptidase A are represented by \leftarrow , and C-terminal amino acids remaining after Edman degradation are represented by \leftarrow . Edman degradation with the Beckman sequencer are identified by \rightarrow and \leftarrow . Tryptic peptides are displayed above the sequence and the other digests below. In order not to complicate the diagram, the various peptides containing allelic substitutions are not shown. The sequence shown in parentheses is ambiguous (see the text).

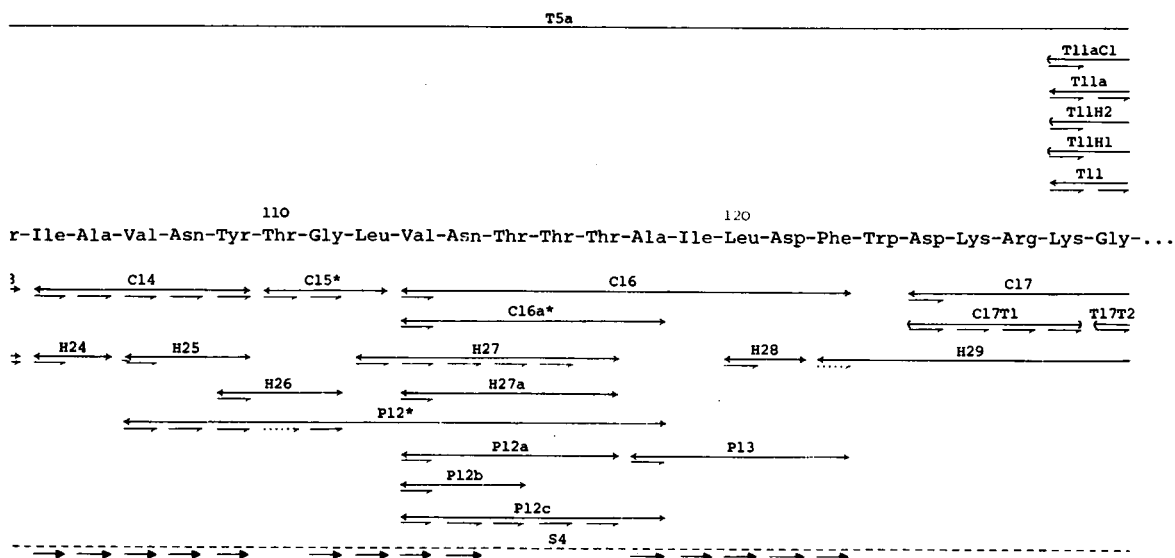
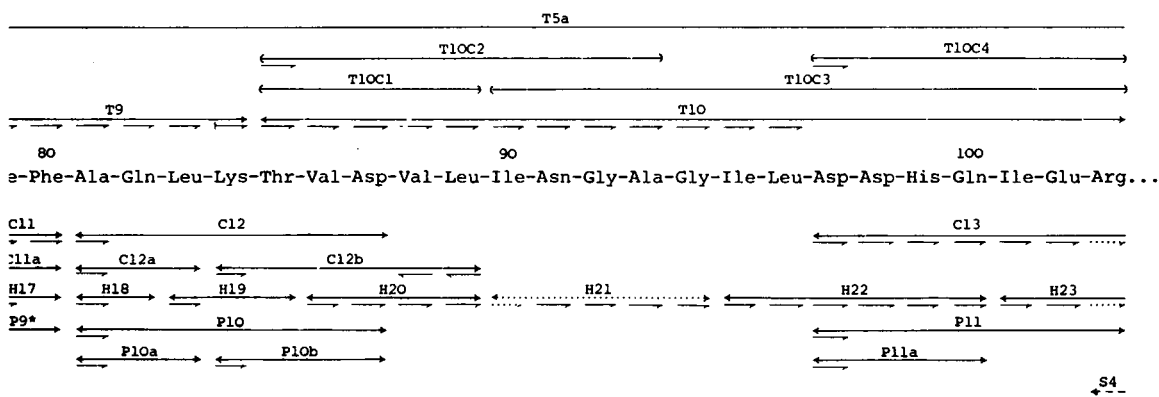
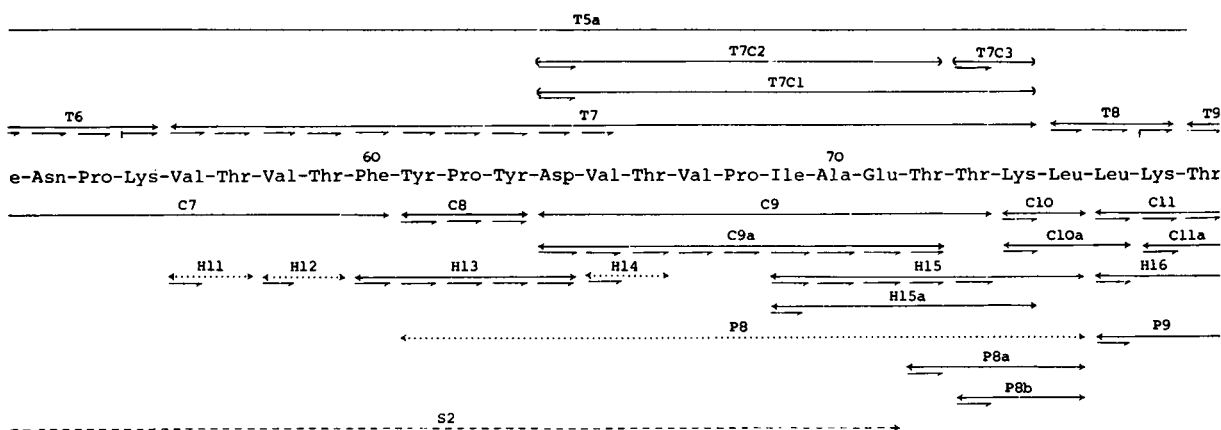


Fig. 3—continued

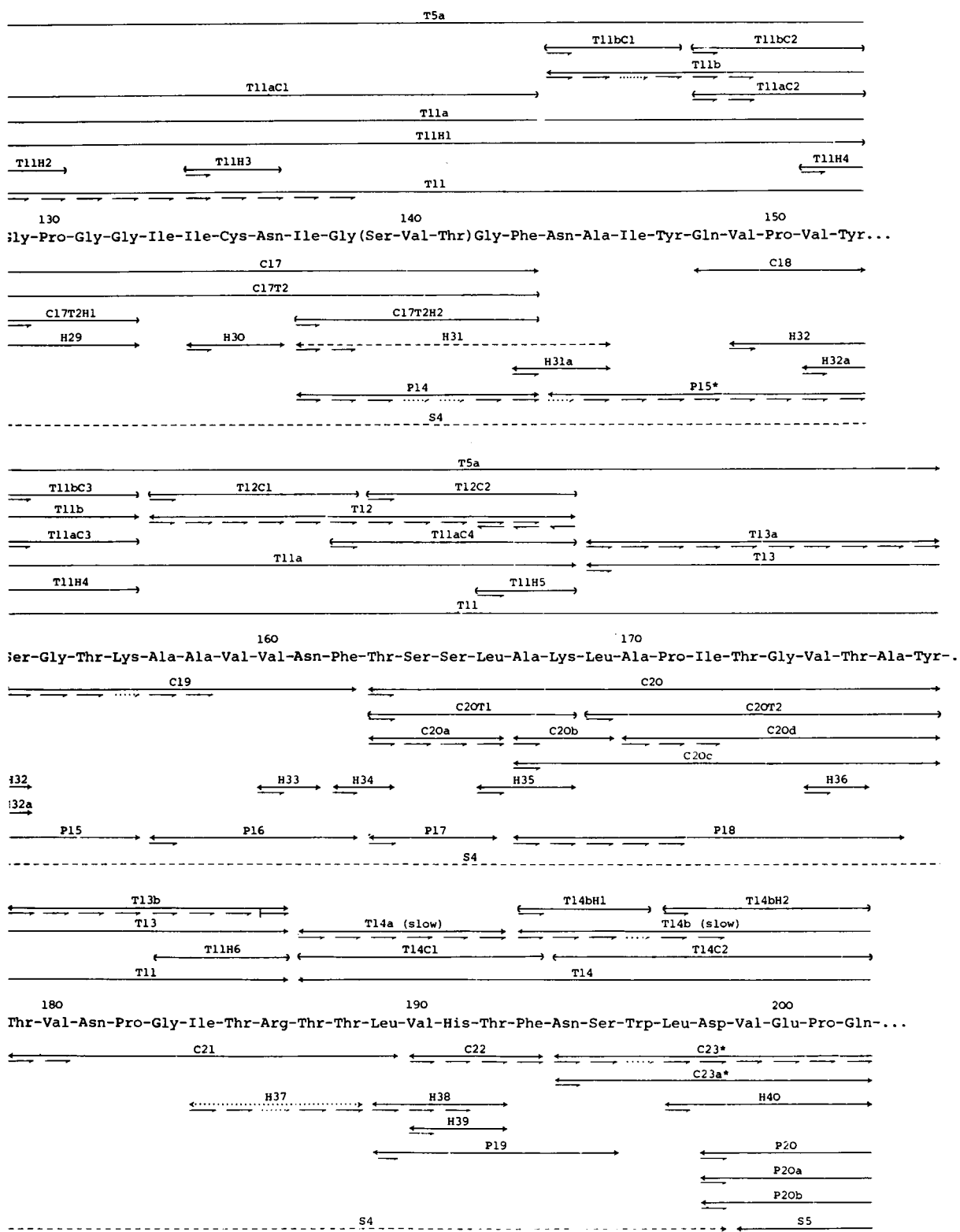


Fig. 3—continued

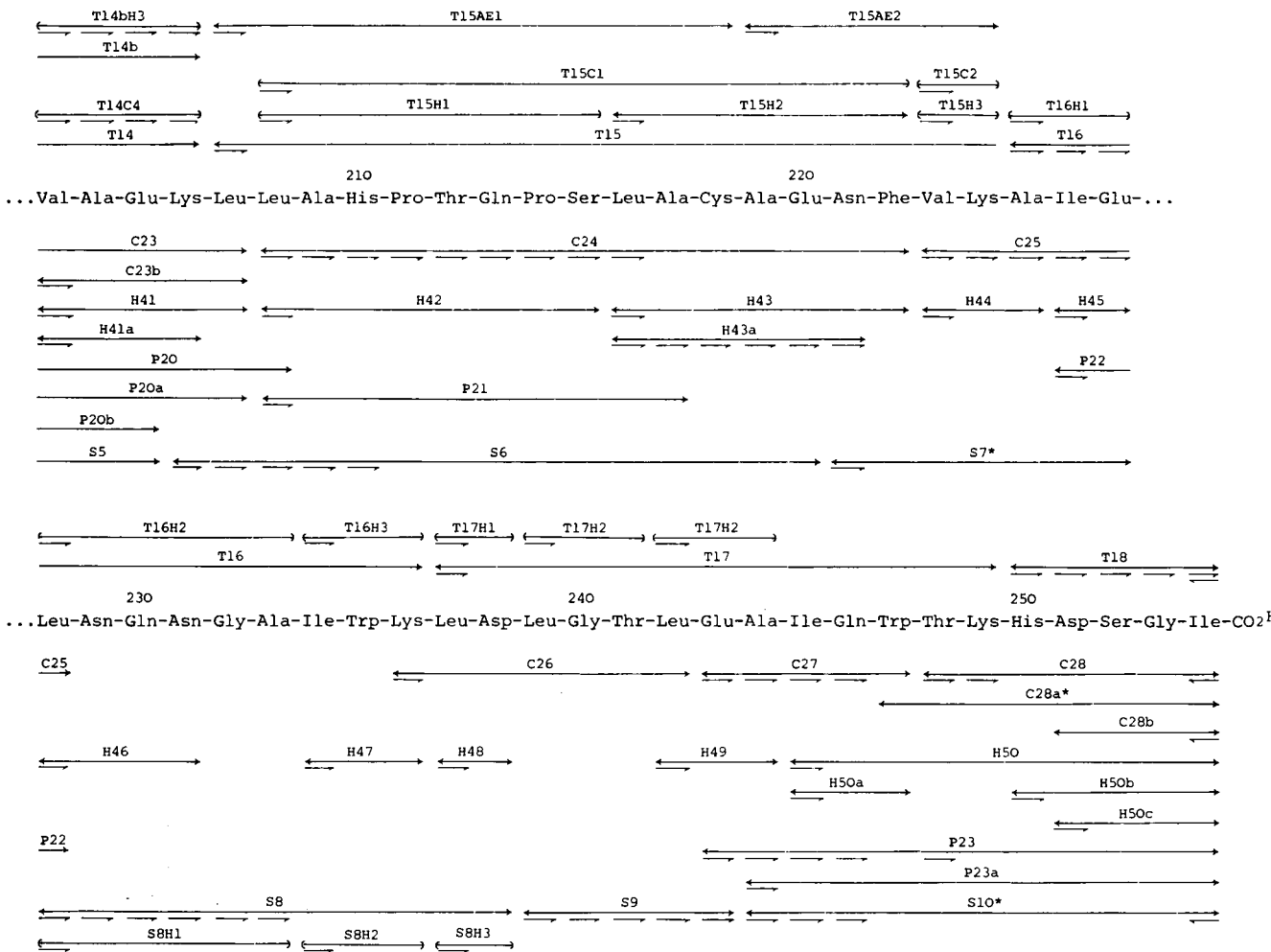


Fig. 3—continued

Thatcher (1977) showed that carboxypeptidase A released isoleucine from the native enzyme molecule. Tryptic peptide T18, chymotryptic peptide C28, thermolysin peptide H50 and staphylococcal-proteinase-V8-digested peptide S10 were all found to have C-terminal isoleucine. Phenylalanine-143 and tyrosine-178 were unusually susceptible to 'pseudo-tryptic' cleavage, being split at a much faster rate than were arginine-39 and lysine-156. Glutamine-25 and -230 were found to be extremely labile. Glutamine-25 was usually isolated as glutamic acid in peptides, with less than 5% of tryptic peptide T2b being recovered in the amidated form. The fact that this was also a major cleavage site for staphylococcal proteinase V8 suggests that deamidation occurs in the native protein and this process may be responsible for the production of multiple elec-

trophoretic forms of the enzyme. Peptide T16 was also detected in two electrophoretic forms, but in this case the interpretation was complicated by the poor solubility properties of this peptide. Schwartz & Jornvall (1976) postulated the presence of an aspartic acid residue at position-229, which must also have been produced by deamidation.

The sequence evidence for the position of residues 138–142 is ambiguous. Both peptide H31 and peptide P14 were impure on quantitative amino-acid analysis, and the dansyl-Edman assignments for this part of the sequence were substandard.

Allelic substitutions

The *Adh*^{N-11} enzyme differed from those produced by *Adh*^{UF} and *Adh*^S alleles by a glycine 15-(*Adh*^{UF} and *Adh*^S)-to-asparagine substitution.

Although this enzyme shows no enzyme activity, Schwartz & Sofer (1976b) showed that *Adh^F/Adh^{N-11}* heterozygotes produced an electrophoretic banding pattern best explained by postulating an ultra-fast electrophoretic mobility for the *Adh^{N-11}* protein. The substitution identified would, if it were in an exposed position on the surface of the molecule, account for the unit change in net charge of the *Adh^F/Adh^{N-11}* hybrid enzyme.

The enzyme produced by *Adh^{UF}* differs from the *Adh^{N-11}* and *Adh^S* enzymes by an alanine-45- (*Adh^{N-11}* and *Adh^S*)-to-asparagine substitution, which is all that is necessary to account for the changed electrophoretic mobility of this electrophoretically ultra-fast alleloenzyme. Peptides T2 and C2 had alanine at position-8 when derived from *Adh^{UF}* protein, but asparagine in the *Adh^{UF}* and *Adh^{N-11}* enzymes.

As described previously, *Adh^S* differed from the *Adh^{UF}* and *Adh^{N-11}* enzymes by a threonine-192-to-lysine (*Adh^F→Adh^S*) substitution (Retzios & Thatcher, 1979; Fletcher *et al.*, 1978).

Discussion

The alcohol dehydrogenase of *Drosophila melanogaster* consists of two identical subunits, each 254 residues in length and of mol.wt. 27400. The enzyme is consequently much smaller in size than the enzyme from horse liver (374 residues/monomer) and yeast (347 residues/monomer (Jornvall *et al.*, 1978). The *Drosophila* enzyme shows no significant similarity in sequence with these two other alcohol dehydrogenases and no homology with mouse dihydrofolate reductase or *Klebsiella aerogenes* ribitol dehydrogenase, which are dehydrogenases of similar low molecular weight (186 and 247 residues respectively) (Dayhoff, 1978). The enzyme from *Drosophila* also contains considerably less cysteine than either horse liver or yeast alcohol dehydrogenase, and the enzyme has no known requirement for a metal cofactor. Secondary-structure prediction by the method of Chou & Fasman (see the Appendix) suggests the presence of a $\beta\alpha\beta\alpha\beta$ super-secondary structural unit (Chothia *et al.*, 1977) in the *N*-terminal half of the molecule. Enzymes that bind nucleotide coenzymes possess a domain invariably constructed from the $\beta\alpha\beta$ unit [the nucleotide-binding or 'Rossmann' fold (Rossmann *et al.*, 1974)]. If the prediction on the *Drosophila* sequence is accurate, and this enzyme is similar in secondary structure to other dehydrogenases, then the nucleotide-binding domain is *N*-terminal, in contrast with the horse liver and yeast enzymes, which have a *C*-terminal nucleotide-binding fold. In the face of such large differences in primary and secondary structure, the conclusion

that the *Drosophila* enzyme is totally unrelated to the horse liver and yeast enzymes is inescapable.

I am immensely grateful for the continued help of Professor B. C. Clarke (University of Nottingham), Dr. M. Ashburner (University of Cambridge) and Mr. A. Retzios, Miss R. Sheik and Dr. R. P. Ambler of this Institution, and for the financial assistance provided by the Medical Research Council and the Science Research Council.

References

- Ambler, R. P. (1973) *Biochem. J.* **135**, 751–758
 Ambler, R. P. & Wynn, M. (1973) *Biochem. J.* **131**, 485–498
 Auffret, A. D., Williams, D. H. & Thatcher, D. R. (1978) *FEBS Lett.* **90**, 324–326
 Chothia, C., Levitt, M. & Richardson, D. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4130–4134
 Clarke, B. C. (1975) *Genetics* **79**, 101–113
 David, J. (1977) *Ann. Biol.* **15**, 9–10
 Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, chapter 6, National Biomedical Research Foundation, Washington
 Drapeau, G. R., Baily, Y. & Doumard, J. (1972) *J. Biol. Chem.* **247**, 6720–6726
 Fletcher, T. S., Ayala, F. J., Thatcher, D. R. & Chambers, G. K. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5609–5612
 Grell, E. H., Jacobson, K. B. & Murphy, J. B. (1968) *Ann. N.Y. Acad. Sci.* **151**, 441–445
 Holder, A. & Ingversen (1978) *Carlsberg Res. Commun.* **43**, 177–184
 Johnson, R. M. & Dennison, C. (1964) *Nature (London)* **204**, 906–907
 Jornvall, H., Eklund, H. & Branden, C.-I. (1978) *J. Biol. Chem.* **253**, 8414–8419
 O'Donnell, J., Mandel, H. C., Krauss, N. & Sofer, W. (1977) *Genetics* **86**, 553–564
 Retzios, A. & Thatcher, D. R. (1979) *Biochimie* **61**, 701–704
 Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) *Nature (London)* **250**, 194–199
 Ryden, A.-C., Ryden, L. & Philipson, L. (1974) *Eur. J. Biochem.* **44**, 105–114
 Schwartz, M. & Jornvall, H. (1976) *Eur. J. Biochem.* **68**, 159–168
 Schwartz, M. & Sofer, W. (1976a) *Nature (London)* **263**, 129–131
 Schwartz, M. & Sofer, W. (1976b) *Genetics* **83**, 125–136
 Thatcher, D. R. (1975) *Biochem. J.* **147**, 313–326
 Thatcher, D. R. (1977) *Biochem. J.* **163**, 317–323
 Thatcher, D. R. & Camfield, R. G. (1977) *Biochem. Soc. Trans.* **5**, 271–272
 Vigue, C. L. & Johnson, F. M. (1973) *Biochem. Genet.* **9**, 213–227
 Woodruff, R. C. & Ashburner, M. (1979a) *Genetics* **92**, 117–132
 Woodruff, R. C. & Ashburner, M. (1979b) *Genetics* **92**, 133–149